UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

November 17, 1999

MASTER FILE

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES #LL-5

MEMORANDUM FOR Michael J. Longini
Chief, Decennial Systems and Contracts Management Office

From: Howard Hogan
Chief, Decennial Statistical Studies Division

Prepared by: Steven P. Hefter
Long Form Sample Design and Estimation Team

Subject: Long Form Sampling Specifications for Census 2000

I. Introduction

The purpose of this memorandum is to document the specifications used for selecting the sample of addresses that will receive long forms in Census 2000. The overall designated long form sampling rate will be approximately 17 percent of all addresses found on the Decennial Master Address File (DMAF) for the United States and Puerto Rico. All addresses in the Island Areas will be enumerated with the long form. The sample design will use target sampling rates of 1-in-2, 1-in-4, 1-in-6, and 1-in-8 to attain the desired, global sampling rate while limiting respondent burden and the disparity in variance between areas. The sampling rates will vary according to the number of addresses in interim census tracts[1] and governmental units, tribal jurisdictions and other specified geographic entities.

In general, the sample of addresses to receive the long form in the mail will be selected from the DMAF. Attachment A provides a flowchart of the general process. Some field operations, where long form sampling will occur, will have preprinted sampling designations in the listing books. Sampling will occur on-site for a subset of service sites, military bases, and maritime vessels. A sample of respondents calling to receive a questionnaire will also be mailed the long form.

[1] An interim census tract is the 1990 census tract or block numbering area (BNA) adjusted to follow those boundaries acceptable to serve as collection block boundaries, such as visible roads and selected legal boundaries. Where the 1990 census tract or BNA boundary was determined to be unacceptable as a collection block boundary, the Geography Division adjusted the boundary to the nearest acceptable feature while maintaining contiguity.

Section II of these specifications contains a list of definitions. Section III lists the files used as inputs to the sampling procedure. Section IV details the sampling frame preparation. Section V provides the procedure for sample selection. Section VI lists the output from the sampling procedure, required for verification.

Any questions regarding these specifications should be directed to Philip Gbur (457-4206, Room 2402A-2) or Steven Hefter (457-4082, Room 2121-2).

II. Definitions

A. Long Form Sampling Entities

Long Form Sampling Entities (LFSEs) are geographic and statistical areas eligible to be considered in assigning the long form sampling rates. The LFSEs are: states, counties, cities, incorporated places (including consolidated cities), minor civil divisions (MCDs) in selected states only[2], American Indian Reservations (AIRs), Tribal Jurisdiction Statistical Areas (TJSAs), Alaska Native Village Statistical Areas (ANVSAs), census designated places (CDPs) in Hawaii only, school districts (based on the 1995-1996 school year) and interim census tracts. American Indian trust lands will follow the designation of their associated reservation.

B. American Community Survey

The American Community Survey (ACS) is a monthly household survey. The ACS questionnaire content is similar to that of the census long form. To minimize respondent burden, the ACS staff will remove the addresses identified on the DMAF that are included in the census long form sample, and exclude them from the ACS sampling frame.

C. Target Universe

The target universe for long form sampling is all eligible addresses in the U.S. and Puerto Rico. All eligible addresses in the Island Areas will be enumerated using the long form. The Island Areas are:

1. U.S. Virgin Islands
2. Guam
3. American Samoa
4. Commonwealth of the Northern Mariana Islands

[2] MCDs in the following states are to be treated as LFSEs: Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and Wisconsin.

D. Sampling Frame

The sampling frames for the target universe consist of:

1. Addresses on the DMAF, which constitutes the foundation of the decennial master address listing control system that is used to control and track census operations;

2. Addresses added during the Update/Leave (U/L) and similar operations;

3. Addresses enumerated as part of the List/Enumerate (L/E) operation;

4. People captured during enumeration of addresses on the Special Place/Group Quarters Control File;

5. People captured during the enumeration of long form eligible service sites, such as soup kitchens and shelters; and

6. Addresses added to the DMAF delivery universe after initial long form sampling before delivery.

E. DMAF Addresses

The criteria for addresses to be included on the DMAF can be found in: DSSD Census 2000 Procedures and Operations Memorandum Series #D-1 "*Specification of the Decennial Master Address File Deliverability Criteria for Census 2000*" from Hogan to Miskura and Marx.

III. Input Files

A. Master Address File (MAF) Extract

The MAF extract contains addresses from the MAF. The Geography (GEO) division will create the MAF extract and deliver it to the Decennial Systems and Contracts Management Office (DSCMO). The criteria for creating the MAF extract is specified in DSSD Census 2000 Procedures and Operations Memorandum Series #D-1 "*Specification of the Decennial Master Address File Deliverability Criteria for Census 2000*" from Hogan to Miskura and Marx.

B. DMAF

The DMAF is the address file used in all Decennial Census operations and includes addresses from the MAF extract . The DMAF is created by adding

various operational codes, counts, and variables to the initial MAF extract. The record layout is under development by the DSCMO.

C. Geographic Reference File - Codes (GRF-C)[3]

The GRF-C is a file containing collection and current geographic codes only for each collection block and may contain multiple records for a block where necessary. See Attachment B for the record layout.

D. Geographic Reference File - Names (GRF-N)

The GRF-N is a file containing names and attributes for geographic entities. See Attachment C for the record layout.

E. School District Equivalency File (SDEQ)

The SDEQ lists all school districts containing each Census 2000 collection block. See Attachment D for the record layout.

F. Accuracy and Coverage Evaluation (A.C.E.) Universe Files

Long form sampling will use the A.C.E. universe files to obtain the 1990 occupied and total HU counts and the 1990 American Indian population and total population counts for each Census 2000 collection block. See Attachment E for the A.C.E. universe file record layout.

Note: When creating the A.C.E. Universe files, maintain the file which includes Remote Alaska for use in long form sampling.

IV. MAF Extract Preparation

A. Identification of Sampling Entities

Although the MAF extract includes all places, the GRF-N identifies which of those are active functioning governmental units. Starting with the MAF extract and using the GRF-N, match on place code. Places having a "Functional Status" of "A", "B", "R", or "T" are considered LFSEs. In addition, places with a "Functional Status" of "S" in Hawaii are also considered LFSEs. Treat any block which is at least partially in a place as if it was entirely within the place.

[3] The record layout for this file, the GRF-N, and the SDEQ are from the Census Bureau's Geography Intranet page at: *http://www.geo.census.gov/tsb/grf/GRFinfo.html*

Note: The "Functional Status" variable has the following possible values:

A active governmental unit;
B active autonomous governmental unit with either a) some overlapping functions with another larger governmental unit or b) shared governmental functions with another governmental unit occupying the exact same territory;
C active entity but with minimal function or administratively subordinate to another governmental unit;
F geographic entity created by the Census Bureau to fill one level of the geographic hierarchy; the geographic entity is properly classified in another level of the hierarchy. An example is the cities in Maine that are also county subdivisions. The Census Bureau creates a record at both the place and county subdivision level for Maine cities; the place level will have the governmental functional status while the county subdivision level will show a functional status of F. These records also are termed "false entities";
I inactive governmental unit;
N nonfunctioning geographic entity (no governmental function or no independent governmental function);
R tribal governmental unit (an American Indian reservation, American Indian trust land entity, or an Alaska Native Village Statistical Area);
S statistical entity; and
T tribal governmental unit (Tribal Jurisdiction Statistical Area)

Using the SDEQ, add to each address record, the codes for all school districts (SDs) that the block, containing the address, is in.

For the remainder of theses specifications, all places with a "Functional Status" code of "A", "B", "R", "T" or "S" (in Hawaii only), *and all SDs* are considered LFSEs.

Create an output variable, *FS*, with the value of the "Functional Status" variable for each LFSE other than SDs. Set *FS* = "G" for each SD.

For every 2000 collection block either wholly or partly in an AIR, TJSA, or ANVSA, set a flag, *ATAFLG* ,with the value of "1" and a value of "0" for all other blocks.

B. Calculation of Measure of Size

Using the A.C.E. universe files and the MAF extract, calculate, for each LFSE and interim census tract, the measure of size (*MOS*) components and the *MOS*.

For the remainder of this memorandum, "collection block" refers to a Census 2000 collection block.

Use the following collection block variables on the A.C.E. universe files. See Attachment E for the complete A.C.E. universe file record layout.

Variable Name	Variable Location	Variable Description
ECOLIR	101-105	Estimated number of American Indian and Alaskan Native renters in the collection block in 1990
ECOLIO	106-110	Estimated number of American Indian and Alaskan Native owners in the collection block in 1990
ECOLHU	161-165	Estimated number of 1990 HUs in the collection block
ECOLOHU	166-170	Estimated number of 1990 occupied HUs in the collection block
ECOLPOP	171-175	Estimated 1990 total non-group quarters (GQ) population in the collection block

1. Compute the following estimated number of 1990 occupied housing units (OHUs) and total HUs. *Do not include the associated trustlands HU count in the following calculations for the AIRs.*

   a. For the i^{th} *LFSE* in the state, calculate $N_{OHU,i}$ and $N_{HU,i}$ as follows:

$$N_{OHU,i} = \sum_{j=1}^{n} ECOLOHU_j$$

$$N_{HU,i} = \sum_{j=1}^{n} ECOLHU_j$$

   where j denotes the individual collection blocks in the i^{th} LFSE and n is the total number of collection blocks in the LFSE.

   If the j^{th} block on the A.C.E. universe file corresponds to m collection blocks, then multiply $ECOLOHU_j$ and $ECOLHU_j$ by $(1/m)$ and use the results in the N_{OHU} and N_{HU} calculations.

   If the j^{th} collection block is split by the LFSE boundary, multiply $ECOLOHU_j$ and $ECOLHU_j$ by 1/2 and use the results in the N_{OHU} and N_{HU} calculations.

b. For the k^{th} interim census tract within the state, calculate $N_{OHU,k}$ and $N_{HU,k}$ as follows:

$$N_{OHU,k} = \sum_{l=1}^{m} ECOLOHU_l$$

$$N_{HU,k} = \sum_{l=1}^{m} ECOLHU_l$$

where l denotes the individual 2000 collection blocks in the k^{th} interim census tract and m is the total number of 2000 collection blocks in the interim census tract.

2. Calculate the following 1990 occupancy rates.

a. For the i^{th} LFSE in the state, calculate $ORATE_i$ as follows:

$$ORATE_i = N_{OHU,i} / N_{HU,i}$$

b. For the k^{th} interim census tract in the state, calculate $ORATE_k$ as follows:

$$ORATE_k = N_{OHU,k} / N_{HU,k}$$

3. Compute the following estimates of the 1990 American Indian population and 1990 total population, $N_{AI,p}$ and $N_{POP,p}$, for the p^{th} AIR, TJSA, or ANVSA[4]. *Do not include the associated trustlands' American Indian population counts in the calculation.*

$$N_{AI,p} = \sum_{r=1}^{q} (ECOLIO_r + ECOLIR_r)$$

$$N_{POP,p} = \sum_{r=1}^{q} ECOLPOP_r$$

where r denotes the individual collection blocks in the p^{th} AIR, TJSA, or ANVSA and q is the total number of 2000 collection blocks in the AIR, TJSA, or ANVSA.

If the r^{th} collection block on the A.C.E. universe file corresponds to m collection blocks, then multiply $ECOLIR_r$, $ECOLIR_r$, and $ECOLPOP_r$ by $(1/m)$ and use the results in the N_{AI} and N_{POP} calculations.

If the r^{th} collection block is split by the AIR, TJSA, or ANVSA boundary, multiply $ECOLIR_r$, $ECOLIR_r$, and $ECOLPOP_r$ by 1/2 and use the results in the N_{AI} and N_{POP} calculations.

[4] If an AIR crosses state boundaries, calculate the *MOS* components and the *MOS* for the parts within each state separately.

4. Compute the following estimate of the 1990 American Indian proportion of total population, $PROPAI_p$, for the p^{th} AIR, TJSA, or ANVSA.

$$PROPAI_p = N_{AI,p} / N_{POP,p}$$

5. Calculate the *MOS* for each LFSE, and interim census tract as follows.

   a. For Mailout/Mailback (MO/MB), U/L, Rural Update/Enumerate, Military, Urban U/L, Urban Update/Enumerate, and MO/MB to U/L conversion areas (Type of Enumeration Areas (TEAs) 1,2,5,6,7,8, and 9):

      i. Compute the *MOS* for the i^{th} LFSE in the state (except AIRs, TJSAs, and ANVSAs) as:

      $$MOS_i = ORATE_i \times MAFHU_i$$

      where "$MAFHU_i$" is the number of addresses on the MAF extract summed across all blocks in the i^{th} LFSE.

      If a collection block is split by the LFSE boundary, add 1/2 of the *MAFHU* for that block in the *MOS* calculation.

      ii. Compute the *MOS* for the k^{th} interim census tract in the state as:

      $$MOS_k = ORATE_k \times MAFHU_k$$

      where "$MAFHU_k$" is the number of addresses on the MAF extract summed across all blocks in the k^{th} interim census tract.

      iii. Compute the *MOS* for the p^{th} AIR, TJSA, or ANVSA in the state as follows. *Do not include the associated trustlands' American Indian MAF extract HU count in the calculation.*

      $$MOS_p = PROPAI_p \times ORATE_p \times MAFHU_p$$

      where "$MAFHU_p$" is the number of addresses on the MAF extract summed across all blocks in the p^{th} AIR, TJSA, or ANVSA.

      If a collection block is split by the LFSE boundary, add 1/2 of the *MAFHU* for that block in the *MOS* calculation.

b. For L/E and Remote Alaska (TEAs 3 and 4)[5]:

i. Define the *MOS* for the i^{th} LFSE in the state (except AIRs, TJSAs, and ANVSAs) as:

$$MOS_i = N_{OHU,i}$$

ii. Define the *MOS* for the k^{th} interim census tract in the state as:

$$MOS_k = N_{OHU,k}$$

iii. Compute the *MOS* for the p^{th} AIR, TJSA, or ANVSA in the state as:

$$MOS_p = PROPAI_p \times N_{OHU,p}$$

C. Assignment of *MOS* to Blocks

Assign each 2000 collection block:

1. A measure of size, *GUMOS,* that corresponds to the smallest *MOS* of all LFSEs in which the block is located.

$$GUMOS = \text{MIN}\{MOS_j\}$$

where $j \in$ of the set of LFSEs containing the block and MIN $\{X_j\}$ is the smallest X_j in the set; and

2. A measure of size, *TRACTMOS,* that corresponds to the MOS_k for the interim census tract in which the block is located.

$$TRACTMOS = MOS_k$$

V. Sample Identification

A. Overview

The primary source of addresses for Census 2000, and hence for long form sampling, will be the DMAF. The take-everys are the reciprocal of the sampling rates.

Addresses added to the DMAF after its initial creation, to be included in the mail-out universe, will be sampled separately using the rates determined in the initial

[5] If any part of a geographic entity is in TEA 3 or 4, treat the entire entity as being in a TEA of 3 or 4 respectively, for purposes of determining the housing unit measure of size.

process. This will allow addresses added to the DMAF frame after initial sample selection to be eligible for long form sampling.

Once the block sampling rate has been determined for all 2000 collection blocks, all adds to the block will be sampled using this rate. *Carry the block sampling rate and follow the sampling pattern in all subsequent sampling operations.* Addresses may be added during the field implementation of the U/L, Rural Update/Enumerate, Urban U/L, and Urban Update/Enumerate operations. A sample of these adds will be selected to receive the long form based on a fixed sampling pattern for the Assignment Area (AA) containing the address.

Addresses will be sampled during the Remote Alaska and L/E operations. Addresses in these AAs will be selected to receive the long form based on a predesignated sampling pattern for the AA containing the address.

All of Puerto Rico, group quarters, and those service sites identified as eligible to be subject to long form sampling (shelters and soup kitchens) will be sampled using a 1-in-6 rate.

B. Assignment to Sampling Strata

Every collection block in the 50 states, the District of Columbia, and Puerto Rico will be assigned to a sampling strata (*SRATE*) and assigned a Target Sampling Rate (*TSR*) based on the block's *GUMOS* or *TRACTMOS*. The *TSRs* for all sampling are 1-in-2, 1-in-4, 1-in-6, and 1-in-8. A block will be assigned to only one sampling strata, and therefore only one *TSR* will be associated with each collection block.

1. Sampling Strata Definitions

The following table defines the four sampling strata and provides other sampling information for them.

Sampling Strata Definition and Associated Sampling Information

SRATE	*TSR*	k	R_k	TE_k	*POS* (rounded)
2	1-in-2	1	R_1	2	0.500
4	1-in-4	2	R_2	4	0.250
6	1-in-6	3	R_3	6	0.167
8	1-in-8	4	R_4	8	0.125

Where:

k represents a sampling stratum defined by the *TSR* ($k = 1, 2, 3, 4$);

R_k represents the random start for the k^{th} sampling stratum;

TE_k represents the take-every for the k^{th} stratum defined as the inverse of the sampling rate; and

POS represents the probability of selection.

2. Sampling Stratum Assignment - General

Assign each block to a sampling stratum as follows:

a. For all blocks in Puerto Rico, set *SRATE* = 6.

b. For all other blocks,

i. If *GUMOS* < 800, set *SRATE* = 2;

ii. If *SRATE* is not equal to 2 and 800 ≤ *GUMOS* < 1200, set *SRATE* = 4;

iii. If *SRATE* is not equal to 2 or 4 *and TRACTMOS* ≥ 2000, set *SRATE* = 8;

iv. Set *SRATE* = 6 for all remaining blocks.

Effectively, with this process, if a LFSE with a *TSR* of 1-in-4 is entirely contained in a second LFSE with a *TSR* of 1-in-6, then all the blocks within the first LFSE will be assigned to the 1-in-4 sampling strata (*SRATE* = 4). Additionally, if a third LFSE with a *TSR* of 1-in-2 should overlap the second LFSE then all blocks in the third LFSE will be assigned a *TSR* of 1-in-2.

3. Sampling Stratum Assignment - Trust Lands

Assign Trust Lands (TL) associated with an AIR to the sampling stratum of the AIR. If the associated AIR has more than one sampling strata assigned to it, assign the lowest sampling rate (largest *SRATE*) to the TL. If the associated AIR is entirely outside the state containing the TL, then set *SRATE* = 2 for the TL.

C. DMAF Frame Address Sampling

1. Basic Procedure

A random start, systematic sampling procedure using variable sampling rates will be used to select the sample from addresses on the initial DMAF for blocks not in L/E or Remote Alaska areas. The sampling procedure will occur independently within each state (including the District of Columbia and Puerto Rico).

a. Select the sample in each sampling stratum, within each state independently. Within each state and sampling stratum, sort the DMAF by county, interim census tract, block, block sequence number[6].

b. Define N_k (k = 1,2,3,4) to be the number of addresses in the k^{th} sampling stratum.

c. For each sampling stratum, generate a sequence of numbers $\{L_j\}$ as follows:

i. Generate a random number, R_k[7], such that $0 < R_k \leq TE_k$;

ii. Let $L_1 = R_k$;

iii. Starting with $j = 2$, calculate $L_j = L_{j-1} + TE_k$ while *CEILING* $[L_j]$[8] $\leq N_k$.

For example, if $R_4 = 3.4$ and $TE_4 = 7.2$ (for the 1-in-8 sampling rate), then $L_1 = 3.4$. The generated L_js would be the sequence: 3.4, 10.6, 17.8, 25.0, 32.2, ...(L_{j-1}+7.2). Therefore, the 4th, 11th, 18th, 25th, 33rd, ..., and the *CEILING* $[L_j]^{th} \leq N_k^{th}$ addresses will be designated to receive a long form.

[6] The block sequence number is to be created by the DSCMO and will reflect an address sort determined by basic street address and apartment number in Mailout/Mailback Areas. The sort in Update/Leave areas will use Mapspot number, basic street address, and apartment number.

[7] Carry all R_k to ten decimal places.

[8] *CEILING* $[X]$ is the nearest integer greater than or equal to X. For example, *CEILING* $[9.87] = 10$.

Note: Generate the four random starts (R_k) once and only once for each state. Continue the sampling pattern, within stratum, across all geographic boundaries, TEAs, and sources of address within state.

d. Create an address level variable *INSAMP*. Within each sampling stratum, set *INSAMP* = 1 for the i^{th} address if the i^{th} address is selected, where i =*CEILING* [L_j] for each L_j. For addresses not selected, set *INSAMP* = 0.

Create an address level variable *ORDER*. For each selected record (*INSAMP* = 1), assign the value of *CEILING* [L_j] within the sampling stratum to the *ORDER* variable. For each record not in sample (*INSAMP* = 0), set *ORDER* = 0.

Assign the probability of selection (*POS*) to each address as given in the sampling strata definitions table above.

Keep a count of the total number of addresses in sample from the k^{th} stratum (n_k).

2. Adds to the DMAF After Initial Sample Selection Before Delivery

As a result of subsequent operations, addresses may be identified for inclusion in the DMAF after the initial long form sample has been selected before questionnaire delivery. Choose a sample of these addresses to receive the long form by continuing the sampling pattern begun for the sampling stratum that the block containing the add(s) is located.

D. Adds in U/L, Rural Update/Enumerate, Urban U/L, and Urban Update/Enumerate Areas

1. Overview

Addresses may be added during the U/L, Rural Update/Enumerate, Urban U/L, and Urban Update/Enumerate field operations. Long form sampling of these adds will be based on the sampling rate for the AA. A single AA may include blocks with different sampling rates. However, we will use one rate for the entire AA. The add pages will be preprinted by the Technologies Management Office (TMO) with the appropriate sampling patterns.

2. Identification of Assignment Area Sampling Rate (*AASR*)

For each AA, define the *AASR* to be the highest *TSR* that occurs within the AA. For example, if an AA includes blocks with *TSR*s of 1-in-4 and 1-in-6 then the *AASR* for that AA would be 1-in-4.

3. Sampling Patterns

Sampling patterns were determined for each sampling strata by randomly determining which unit would be designated to receive the long form. Use the appropriate sampling patterns for each *AASR* as follows (S = short form and L = long form):

AASR	Sampling Pattern	*POS*
1-in-2	S L S L ...	0.500
1-in-4	S L S S S L S S ...	0.250
1-in-6	S S L S S S S S L S S S ...	0.167
1-in-8	S L S S S S S S S L S S S S S S ...	0.125

E. L/E and Remote Alaska (TEAs 3 and 4)

1. Overview

A portion of the population will be enumerated in Census 2000 using the L/E methodology, which is a field operation. The address listing books used in L/E areas and Remote Alaska will be preprinted with the appropriate sampling patterns. Long form sampling of these addresses will be based on the sampling rate for the AA. Two sampling rates, 1-in-2 and 1-in-6, will be used in L/E and Remote Alaska AAs. A single AA may include blocks with different *TSR*s, however, we will use only one rate for the entire AA.

2. Identification of Assignment Area Sampling Rate (*AASR*)

For each AA, determine the AA's sampling rate as follows.

a. Determine the highest *TSR* within the AA.

b. Use the highest *TSR* within the AA in conjunction with the following table to determine the AA's sampling rate. Sample all blocks within the AA using the *AASR*.

Sampling Rate Assignment for L/E and Remote Alaska AAs	
Highest *TSR* Within AA	*AASR*
1-in-2	1-in-2
1-in-4	
1-in-6	1-in-6
1-in-8	

For example, if an AA includes blocks with *TSR*s of 1-in-4 and 1-in-6 then the highest block *TSR* within the AA is 1-in-4. Referring to the table, the *AASR* would be 1-in-2.

3. Sampling Patterns

The sampling patterns will be preprinted for both the 1-in-2 and 1-in-6 sampling rates. The sampling pattern will be randomized by selecting a random starting point for each address listing book.

4. Randomization Procedure for L/E and Remote Alaska Address Registers

As with all systematic samples, the starting point for the address registers in L/E and Remote Alaska AAs must be randomized. Define *RS* to be the number of lines to cross out on the first page of an AA's address register.

For each L/E and Remote Alaska AA, perform the following.

a. Generate a random number *RN*, such that $0 < RN < 1$.

b. For 1-in-2 AAs

i. If $RN < 0.5$, set $RS = 0$. No action is needed.

ii. If $RN \geq 0.5$, set $RS = 1$.

c. For 1-in-6 AAs

i. If $RN < 0.1667$, set $RS = 0$. No action is needed.

ii. If $0.1667 \leq RN < 0.3333$, set $RS = 1$.

iii. If $0.3333 \leq RN < 0.5000$, set $RS = 2$.

iv. If $0.5000 \leq RN < 0.6667$, set $RS = 3$.

v. If $0.6667 \leq RN < 0.8333$, set $RS = 4$.

vi. If $0.8333 \leq RN < 1.0000$, set $RS = 5$.

For each L/E and Remote Alaska AA, deliver the *RS* and the *AASR* to the TMO.

F. Group Quarters

GQ enumeration will occur at addresses identified in the Special Place/Group Quarters Control File. The sampling rate for people residing in GQs will be 1-in-6 resulting in a probability of selection of approximately 0.167. Using the same sampling pattern as was determined for Update/Leave adds, select people to receive a long form based on the following sampling pattern, (S = short form and L = long form):

S S L S S S S S L S S S ...

Any added housing units identified during the Advance Listing will be listed on U/L add pages which will have the sample designation with the fixed sampling pattern as described above in Section V.D.

G. Service Sites

Shelters and soup kitchens, which are part of the Service Based Enumeration (SBE) operation, will include a sample of long forms. The sampling rate for these locations will be 1-in-6 resulting in an approximate probability of selection of 0.167. As there is no listing prior to the enumeration, the enumerators will have a stack of questionnaires with long forms interspersed among the short forms and distribute them in the order of the stack. Using the same sampling pattern as determined for U/L adds, use the following pattern in creating the stacks of forms (S = short form and L = long form):

S S L S S S S S L S S S ...

H. Telephone Questionnaire Assistance

1. Overview

The Telephone Questionnaire Assistance (TQA) operation will take requests for questionnaires to be mailed and will take short form interviews only. Individuals who telephone to request a questionnaire will either receive their designated form type or be subject to a 1-in-6 long form sampling rate depending upon whether they have their census

identification number (ID). Individuals who call to provide an interview on the telephone are not eligible for long form sampling.

2. Identification of Appropriate Form Type

   a. If the caller can provide the TQA staff their census ID, the address will be located on the DMAF and mailed the appropriate form type (short or long).

   b. If the caller either does not have a census ID or cannot identify their census ID, a 1-in-6 sample of these requests will be mailed a long form. Using the same systematic sampling pattern used for U/L adds, use the following pattern to determine the appropriate form type (S = short form and L = long form):

      S S L S S S S S L S S S ...

I. Addresses Added by the United States Postal Service

Addresses added by the United States Postal Service (USPS) since the November Delivery Sequence File will be included in a February USPS updated added address file delivery. The new addresses which are geocoded in time will be included in the February 28, 2000 MAF Extract Delivery. Sample these adds as specified in Section V.C.2.

J. Military Bases and Maritime Vessels

People living in housing units at military installations will receive questionnaires by mail and be sampled using the procedures described in Section V.C. Sample others (such as those living in barracks on base and those on vessels) that are enumerated at their work stations via Military Census Reports or Shipboard Census Reports as appropriate, using a fixed 1-in-6 sampling rate. Base the sampling on the last four digits of the respondent's Social Security Number (SSN).

Identify persons to complete the long form by determining the last four digits of their SSN. If the last four digits are greater than 8332, instruct the person to complete the sample portion of the questionnaire.

VI. Output

Provide the Decennial Statistical Studies Division (DSSD) each of the following outputs after initial sample selection *and File D each time the procedures specified in Section V.C are applied.* Create files A→D as described in the following corresponding sections, for each state (including Washington, DC) and Puerto Rico.

A. Provide a block level file with the results of the sampling. A suggested record layout is provided in Attachment F.

B. Provide a LFSE level output file to the DSSD. A suggested record layout is provided in Attachment G.

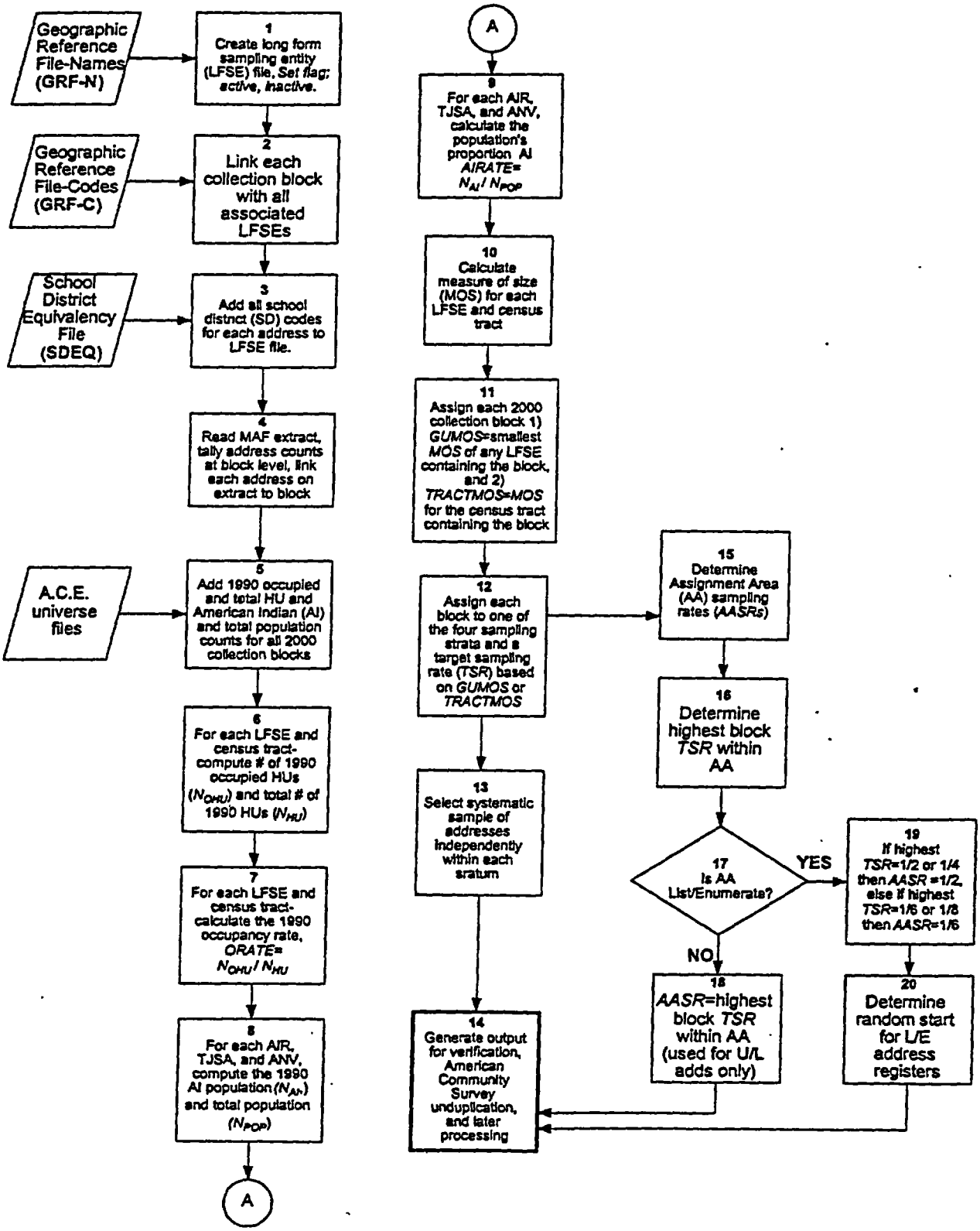
C. Provide an interim census tract level output file to the DSSD. Attachment H contains a suggested record layout.

D. To allow the ACS to unduplicate it's sample with the initially selected portion of the Census 2000 long form sample, and for the DSSD verification, create fifty-two files (one per state, Washington, DC and Puerto Rico), that list the long form sampling universe and the addresses designated to be in sample (*INSAMP* = 1). These files (with the exception of Puerto Rico) are to be delivered to the ACS staff upon notification by the DSSD that the long form sample has been approved. Attachment I contains the requested variables and their descriptions. Please provide the record layout, which is left to the discretion of the DSCMO, to the DSSD and the DSMD (Scot Dahl).

cc:

DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
Statistical Design Team Leaders
Long Form Sample Design and Estimation Team Distribution List
M. Hudson (DSSD)
R. Feldpausch
G. Leithauser (FLD)
C. Moore
K. Medina
L. Ramirez (GEO)
A. Smith (POP)
D. Smith
E. Gore (DMD)
A. Quinlan

Census 2000 Long Form Sampling Procedure



Record Layout for Collection GRF-C

Variable	Location
Collection FIPS State Code	1:2
Collection County FIPS Code	3:5
Collection Block Number	6:10
Collection Block Suffix	11:11
TEA	12:12
LCO	13:16
Address Listing CFO/ Block Canvass FOS	17:20
Address Listing FOS/ Last 2 Digits of BC FOS	21:22
Crew Leader District	23:24
Operation Code	25:26
Assignment Area Code	27:30
Pseudo Tract Code	31:36
Current FIPS State Code	37:38
Current FIPS County Code	39:41
Current MCD/CCD Code	42:44
MCD Flag	45:45
Current Place Code	46:49
Place Flag	50:50
AIR/ANVSA/TDSA/TJSA	51:54
AIR Flag	55:55
American Indian Trust Land Flag	56:56
Consolidated City	57:57
Consolidated City Flag	58:58
Land Area	59:63
Estimated Population Count	64:68
Estimated Housing Count	69:73
Block Type Flag	74:74
First Zip Code	75:79
Cluster Number	80:84
Cluster Check Digit Number	85:85
TEA Recode for ICM	86:86
A.C.E. Sample Flag	87:87

Record Layout for GRF-N

Variable	Location
Record Type	1:2
FIPS State Code	3:4
FIPS County Code	5:7
County Subdivision	8:10
Sub-Minor Civil Division	11:12
Place Code	13:16
Consolidated City	17:20
AI/ANA/HH	21:24
American Indian Sub-Reservation	25:27
American Indian Trust Land Indicator	28:28
AI/ANA/TJSA/TDSA/HH Entity Type	29:29
American Indian Reservation Type	30:30
Alaska Native Regional Corporation	31:32
CMSA	33:34
FIPS MSA/CMSA	35:38
FIPS PMSA	39:42
Region	43:43
Division	44:44
105th/106th Congressional District	45:46
Voting District	47:52
State Legislative District - Lower Chamber	53:55
State Legislative District - Upper Chamber	56:58
Reserved Space	59:61
Urbanized Area	62:66
Oregon Urban Growth Area	67:71
Enterprise Zone/Empowerment Community	72:77
FIPS 55 Code for Entity	78:82
FIPS 55 Class Code for Entity	83:84
State/US Abbreviation	85:86
Functional Status	87:87
Legal/Statistical Area Description Code	88:89
Entity Place Description Code	90:90
Place Description Code (MCD)	91:91
Place Description Code (County)	92:92
Voting District Indicator	93:93
Geographic Change User Note Indicator	94:94
Area Name (w/out LSAD)	95:95

Record Layout for Census 2000 School District Equivalency File (SDEQ)

Variable	Location
Collection FIPS State Code	1:2
Collection FIPS County Code	3:5
Collection Block Number	6:10
Collection Block Suffix	11:11
Collection Block Split Indicator	12:12
Land/Water Flag	13:13
Reserved Space	14:14
School District Code	15:19
School District Level	20:20
School District Code	21:25
School District Level	26:26
School District Code	27:31
School District Level	32:32
School District Code	33:37
School District Level	38:38
School District Code	39:43
School District Level	44:44
School District Code	45:49
School District Level	50:50
School District Code	51:55
School District Level	56:56
School District Code	57:61
School District Level	62:62
School District Code	63:67
School District Level	68:68
School District Code	69:73
School District Level	74:74

A.C.E. Universe File Record Layout

Variable Description	Name	Location
State	STATE	1-2
County	COUNTY	3-5
Interim Tract (a.k.a. pseudo-tract)	ITRACT	6-11
Block Number	COLBLOCK	12-16
Blank		17-17
Cluster Number (geography not A.C.E.)	GCLUS	18-22
Blank		23-23
Cluster Size code	CLUSSIZE	24-24
1 = Clusters with 0 HUs		
2 = Clusters with 1 HUs		
3 = Clusters with 2 HUs		
4 = Clusters with between 3 and 5 HUs		
5 = Clusters with between 6 and 9 HUs		
6 = Clusters with between 10 and 19 HUs		
7 = Clusters with between 20 and 29 HUs		
8 = Clusters with between 30 and 79 HUs		
9 = Clusters with 80 or more HUs		
Blank		25-25
Block Area (Sq. Miles)	BAREA	26-33
Blank		34-34
Block Perimeter (Miles)	BPERIM	35-40
Blank		41-41
Block Cluster Area (Sq. Miles)	BCAREA	42-49
Blank		50-50
Block Cluster Perimeter (Miles)	BCPERIM	51-56
Number of HUs in cluster	NHU	57-61
Number of HUs in block	NHUBLOCK	62-66
Block TEA	TEA	67-67
1 = Mailout/Mailback		
2 = Update/Leave		
3 = List/Enumerate		
5 = Rural Update/Enumerate		
6 = Military		
7 = Urban Update/Leave		
8 = Update/Leave to Mailout/Mailback conversions		
9 = Mailout/Mailback to Update/Leave conversions		

TEA Group for Block Cluster	TEABC	68-68
A= Mailout/Mailback or Urban Update/Leave or Update/Leave to Mailout/Mailback conversions		
B= Update/Leave or Rural Update/Enumerate		
C=List/Enumerate		
D=Military		
E=Mailout/Mailback to Update/Leave conversions		
2000 MAF HUs count	NHUM	69-73
' ' Blank if no HU count available		
1990 ACF HUs count	NHU90	74-78
' ' Blank if no HUs count available		
Housing Unit Count Indicator	HUIND	79-79
1 = from 2000 MAF		
2 = from 1990 ACF		
Invisible Boundary Collapse Indicator	INV	80-80
0 = No		
1 = Yes (Collapsing across Invisible Boundary in BC)		
American Indian Country Indicator	AICIND	81-81
0 = No American Indian Country		
1 = American Indian Reservation/trust land		
2 = Tribal jurisdiction statistical area/ Alaska Native Village statistical area/ tribal designated statistical area		
Military Indicator	MILIND	82-82
0 = No Military Area		
1 = Block contains Military Area		
Collapsed Enclosed Block Indicator	CEBI	83-83
0 = Otherwise		
1 = An enclosed block has been forced to collapse		

Blank		84-90
2000 Collection Block Estimated Number of:		
Hawaiian and Pacific Islander Renter	ECOLPIR	91-95
Hawaiian and Pacific Islander Owner	ECOLPIO	96-100
American Indian and Alaska Native Renter	***ECOLIR***	***101-105***
American Indian and Alaska Native Owner	***ECOLIO***	***106-110***
Asian Renter	ECOLAR	111-115
Asian Owner	ECOLAO	116-120

Hispanic Renter	ECOLHR	121-125
Hispanic Owner	ECOLHO	126-130
Black Renter	ECOLBR	131-135
Black Owner	ECOLBO	136-140
White and Other Renter	ECOLOR	141-145
White and Other Owner	ECOLOO	146-150
Total Renters	ECOLR	151-155
Total Owners	ECOLO	156-160
Total Housing Units	***ECOLHU***	***161-165***
Occupied Housing Units	***ECOLOHU***	***166-170***
Total People (Non-GQ)	***ECOLPOP***	***171-175***
Estimated 1990 urbanicity of the 2000 collection block	ECOLURB	176-176
1 = Urban Area with 1990 population ≥ 250,000		
2 = Other Urban Area		
3 = Non-Urban Area		
Blank		177-180
2000 Collection Block Cluster Estimated Number of:		
Hawaiian and Pacific Islander Renter	ECLUSPIR	181-185
Hawaiian and Pacific Islander Owner	ECLUSPIO	186-190
American Indian and Alaska Native Renter	ECLUSIR	191-195
American Indian and Alaska Native Owner	ECLUSIO	196-200
Asian Renter	ECLUSAR	201-205
Asian Owner	ECLUSAO	206-210
Hispanic Renter	ECLUSHR	211-215
Hispanic Owner	ECLUSHO	216-220
Black Renter	ECLUSBR	221-225
Black Owner	ECLUSBO	226-230
White and Other Renter	ECLUSOR	231-235
White and Other Owner	ECLUSOO	236-240
Total Renters	ECLUSR	241-245
Total Owners	ECLUSO	246-250
Total Housing Units	ECLUSHU	251-255
Occupied Housing Units	ECLUSOHU	256-260
Total People (Non-GQ)	ECLUSPOP	261-265
Blank		266-275
Estimated 1990 urbanicity of 2000 block cluster	ECLUSURB	276-276
1 = Urban Area with 1990 population ≥ 250,000		
2 = Other Urban Area		
3 = Non-Urban Area		
Size Category	SIZECAT	277-277
1 = Small (0-2 HUs)		

2 = Medium (3-79 HUs)		
3 = Large (80+ HUs)		
Number of sampling strata in state	NSSINST	278-278
Sample stratum	SS	279-279
1 = Small		
2 = Medium (non-AIR)		
3 = Large (non-AIR)		
4 = American Indian Reservation		
Blank		280-285
2000 Collection Block Cluster Proportion of Population that is:		
Hawaiian and Pacific Islander Renter	CLUPPIR	286-290
Hawaiian and Pacific Islander Owner	CLUPPIO	291-295
American Indian and Alaska Native Renter	CLUPIR	296-300
American Indian and Alaska Native Owner	CLUPIO	301-305
Asian Renter	CLUPAR	306-310
Asian Owner	CLUPAO	311-315
Hispanic Renter	CLUPHR	316-320
Hispanic Owner	CLUPHO	321-325
Black Renter	CLUPBR	326-330
Black Owner	CLUPBO	331-335
White and Other Renter	CLUPOR	336-340
White and Other Owner	CLUPOO	341-345
Renters	CLUPR	346-350
Owners	CLUPO	351-355
Blank		356-364
Demographic/Tenure group (code)	DTCODE	365-366
Demographic/Tenure group (label)	DTLABEL	367-368
Region	REGION	369-369
Division	DIV	370-370

Suggested Record Layout for Block Level Output File

Variable	Description	Location
STATE	State	1-2
COUNTY	County	3-5
BLOCK	Collection Block Number	6-11
TEA	Type of Enumeration Area	13
LCO	Local Census Office Code	15-18
AA	Assignment Area Code	20-25
MAFHUS	Number of HUs from MAF extract	27-31
ACEHUS	Number of HUs from A.C.E. Universe file	32-36
ACEOHUS	Number of OHUs from A.C.E. Universe file	37-41
ACEPERS	Number of persons from A.C.E. Universe file	42-46
ACEAIPER	Number of AI persons from A.C.E. Universe file	47-51
TOTLFS	Total number of HUs selected to receive long form	52-56
LFSEMOS	The LFSE Measure of size for the block	57-65
TRMOS	The measure of size for the interim census tract	66-74
BSAM	Target Sampling Rate	75-76
AASR	The Assignment Area sampling rate	77-78
TRACT	Interim census tract code	80-85
BLKPRT1	LFSE1 split block indicator	92
CODE1	LFSE1 code	94-100
MOS1	The measure of size for LFSE1	101-109
BLKPRT2	...Repeat Decriptions for LFSEs 2-14	110
CODE2		112-118
MOS2		119-127
BLKPRT3		128
CODE3		130-136
MOS3		137-145
BLKPRT4		146
CODE4		148-154
MOS4		155-163
BLKPRT5		164
CODE5		166-172
MOS5		173-181
BLKPRT6		182
CODE6		184-190
MOS6		191-199

Variable	Description	Location
BLKPRT7		200
CODE7		202-208
MOS7		209-217
BLKPRT8		218
CODE8		220-226
MOS8		227-235
BLKPRT9		236
CODE9		238-244
MOS9		245-253
BLKPRT10		254
CODE10		256-262
MOS10		263-271
BLKPRT11		272
CODE11		274-280
MOS11		281-289
BLKPRT12		290
CODE12		292-298
MOS12		299-307
BLKPRT13		308
CODE13		310-316
MOS13		317-325
BLKPRT14		326
CODE14		328-334
MOS14		335-343

Suggested Record Layout for LFSE Output Files

Variable	Description	Location
CODE	SD, GU, or AIR code	1-7
FUNCSTAT	Functional status (from GRF-C)	11-11
BLOCKS	Number of blocks in LFSE	13-18
LEBLOCKS	Number of List/Enumerate blocks	19-24
MOS	The measure of size	25-33
MAFHUS	Number of HUs from the MAF extract	34-42
ACEHUS	Number of HUs from the A.C.E. Universe file	43-51
ACEOHUS	Number of OHUs from the A.C.E. Universe file	52-60
ACEPERS	Number of persons from the A.C.E. Universe file	61-69
ACEAIPER	Number of American Indians from the A.C.E. Universe file	70-78
TOTLFS	Total number of long forms chosen	79-87
HUSIN2	Number of HUs in 1-in-2 sampling strata	88-96
HUSIN4	Number of HUs in 1-in-4 sampling strata	97-105
HUSIN6	Number of HUs in 1-in-6 sampling strata	106-114
HUSIN8	Number of HUs in 1-in-8 sampling strata	115-123
LFSIN2	Number of long forms chosen in the 1-in-2 sampling strata	124-132
LFSIN4	Number of long forms chosen in the 1-in-4 sampling strata	133-141
LFSIN6	Number of long forms chosen in the 1-in-6 sampling strata	142-150
LFSIN8	Number of long forms chosen in the 1-in-8 sampling strata	151-159
ORATE	Occupancy rate as calculated from the A.C.E. Universe file	160-165
PROPAI	Proportion of American Indians from A.C.E. Universe file	166-171

Suggested Record Layout for Interim Census Tract Output Files

Variable	Description	Location
TRCODE	Interim census tract code	1-9
BLOCKS	Number of blocks	13-18
LEBLOCKS	Number of List/Enumerate blocks	19-24
MOS	Measure of size	25-33
MAFHUS	Number of HUs from the MAF extract	34-42
ACEHUS	Number of HUs from the A.C.E. Universe file	43-51
ACEOHUS	Number of OHUs from the A.C.E. Universe file	52-60
ACEPERS	Number of persons from the A.C.E. Universe file	61-69
ACEAIPER	Number of American Indians from the A.C.E. Universe file	70-78
TOTLFS	Total number of long forms chosen	79-87
HUSIN2	Number of HUs in the 1-in-2 sampling strata	88-96
HUSIN4	Number of HUs in the 1-in-4 sampling strata	97-105
HUSIN6	Number of HUs in the 1-in-6 sampling strata	106-114
HUSIN8	Number of HUs in the 1-in-8 sampling strata	115-123
LFSIN2	Number of long forms chosen in the 1-in-2 sampling strata	124-132
LFSIN4	Number of long forms chosen in the 1-in-4 sampling strata	133-141
LFSIN6	Number of long forms chosen in the 1-in-6 sampling strata	142-150
LFSIN8	Number of long forms chosen in the 1-in-8 sampling strata	151-159
ORATE	Occupancy rate from the A.C.E. Universe file	160-165
PROPAI	Proportion of American Indians from A.C.E. Universe file	166-171

Requested Variables for the ACS/Long Form Sample File

The following variables are to be included in the requested set of files for delivery to the ACS staff. This file will also be used by the DSSD for verification. Each record on the file(s) is an address from the DMAF.

Variable Name	Variable Description
ST	State code (including the District of Columbia)
COU	County code
TRACT	Tract code
BLK	2000 collection block code
BSAM	Sampling Stratum Indicator 1 = 1-in-2 (k = 1) 2 = 1-in-4 (k = 2) 3 = 1-in-6 (k = 3) 4 = 1-in-8 (k = 4)
INSAMP	Sample Indicator 1 = selected for long form 0 = not selected for long form
MAFID	Master Address File ID
TE	The take-every used to select the record
RANDST	Random start
ORDER	For records in sample, *ORDER* = *CEILING* [L_j] For records not in sample, *ORDER* = 0
POS	Probability of Selection
LCO	Local Census Office code
LJ	The ceiling of the sampling algorithm sequence number
STRCOUNT	Count of the number of long forms, so far, within strata